DOCUMENT RESUME

ED 208 941 IR 009 750

AUTHOR Mischo, William H.
TITLE Study and Development of Mechanized Indexing Procedures That Can Be Applied to the Production of Subject-Enhanced Keyword Indexes. Final Report.
INSTITUTION Iowa State Univ. of Science and Technology, Ames. Library.
SPONS AGENCY Council on Library Resources, Inc., Washington, D.C.
PUB DATE Aug 79
GRANT CLR-677
NOTE 34p.; Appendices may not reproduce.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Automatic Indexing; College Libraries; *Computer Programs; Higher Education; *Library Catalogs; Library Collections; Online Systems; *Permuted Indexes; *Reference Materials; *Subject Index Terms
IDENTIFIERS Iowa State University

ABSTRACT

This paper describes a computer-assisted printed index which provides expanded subject access to the more than 6000 titles in the Iowa State University Library reference collection. Designed to complement existing catalogs with a minimum of duplication of information, the index displays abbreviated length records. The limitations of subject access to reference materials via Library of Congress Subject Headings are discussed, and a system for assigning subject descriptors to reference titles is described. Central to the project are custom designed software routines that expedite data entry, minimize keyboarding, and use a combinatorial-based method to produce multiple entry points from single input strings. Included as appendices are an example page from the reference collection index, a list of form and content designators, sample catalog card revisions, and the program for a keyword out of context (KWOC) index. There are 21 references. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208841

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Final Report

Council on Library Resources Grant CLR 677 to Iowa State University for the Study and Development of Mechanized Indexing Procedures that can be Applied to the Production of Subject-Enhanced Keyword Indexes

IOWA STATE UNIVERSITY

Ames, Iowa

BEST COPY AVAILABLE

August 1979

William H. Mischo
Principal Investigator
Iowa State University Library

BEST COPY AVAILABLE

IR009752

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William H. Mischo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Library automation activities directed to reference and information department services have principally involved on-line systems designed to enhance retrieval of bibliographic citations contained in secondary abstracting and indexing (A & I) services. Little attempt has been made to expand access to information contained in other types of reference collection materials--such as dictionaries, handbooks, bibliographies, etc. Several studies have investigated the feasibility of automating search procedures to biographical reference books[1-2] or other discrete subsets of a reference collection.[3-4] It should be noted that these studies were performed in laboratory environments and did not involve an operational system in a library setting.

Catalog use studies have shown that different library users have different needs.[5] Reference collection materials are approached by users who are, in most cases, interested in items that can fill a specific informational need, rather than with locating known-item author or title citations. Patrons typically will be seeking a reference tool that can provide either an on-the-spot piece of information (e.g., an address, a definition or statistic, a citation verification, etc.) or a bibliography of primary source citations in a specific subject area. These user needs demand a system capable of providing in-depth access to specific types of reference tools (e.g., handbooks, A & I services, dictionaries) associated with specific subject areas (e.g., archaeology, nuclear engineering). This is information retrieval one step removed from the typical machine-assisted reference service in that what is desired is a list of candidate reference titles for meeting the information need. In the case of A & I services, for example, the patron is interested in identifying from among the myriad A & I services those containing citations in a subject area (e.g., archaeology).

However, the standard source of subject access to reference materials in college and research libraries--Library of Congress Subject Headings (LCSH)--is unable to completely meet these user needs. The inadequacies of subject access through LCSH[6-11] (outmoded descriptor terms, inconsistent forms of headings, lack of term specificity, lack of indexing exhaustivity, etc.) are accentuated by the paramount importance of the subject approach to reference collection materials.

Enhanced access using terms taken from the titles of reference works can be easily shown to be of limited effectiveness. The title-derivative or KWIC/KWOC approach has serious retrieval limitations.[12] Thus, the information contained in the machine-readable network records or MARC related products is unable to provide the in-depth subject access terms required for reference titles, and the need must be met by local augmentation.

Although augmentation of monographic titles for retrieval in an on-line environment has been successfully demonstrated by Atherton,[13] the nature of reference requests--that is, for a type of tool covering a subject area--suggests that a Boolean search capability in the retrieval of reference titles is not necessary. Rather, a printed format index, featuring exhaustive subject analysis and allowing the identification of the form of the reference tool, is capable of effectively providing the needed access. Malinconico[14] has suggested that increasing the number of precoordinated subject access points in a printed index will tend to duplicate the document retrieval recall performance of an interactive system. This paper reports on the design and construction of a computer-produced printed index providing the needed expanded access to titles in the Iowa State University Library Reference Collection.

Objectives

This project focused on the development of mechanized indexing procedures that can be applied in the production of printed indexes designed to improve access to materials in academic library reference collections. Particular emphasis was placed on:

1. The identification of optimum access points for reference titles and the design of an indexing language--specifically testing the efficacy of an uncontrolled vocabulary index;
2. The development of streamlined data preparation and data entry procedures;
3. Testing the feasibility of a printed format index and attempting to determine the number of data elements needed in index entries;

4. In conjunction with 1 and 2, the development of custom-designed software utilizing the computer's facility in manipulating word strings to produce index entries.

Procedures

The Iowa State University Library Reference Collection is a multidisciplinary, centralized collection of over 6,000 titles. Initially, titles from the Reference Collection were examined and prototype indexes produced and studied for form and type of entry points and index format. It was decided that the printed index would display an abbreviated bibliographic record consisting of lead term, title, publication date and call number. Palmer[15] has suggested that a short entry catalog with the above-mentioned fields plus author's name will function effectively in an academic library setting. However, in this application the author is a less frequently used access point and a less important element for identifying the work. The index functions as a finding aid, as opposed to a source of full bibliographic information, and complements the Library's other bibliographic files (Card and Serials Catalog) with a minimum of duplication of information.

A hybrid indexing language was devised, with entry points for individual titles to include: (1) specific descriptor terms taken from the table of contents and body of the work, (2) title entries, (3) significant words and phrases appearing in the title or subtitle, (4) author, compiler or editor on an infrequent basis, and (5) form and content designations, identifying the type of reference material, e.g., encyclopedia, atlas, biographical information, etc. Title and personal name access points, although secondary in importance to exhaustive subject analysis, were judged to be approached frequently enough for inclusion as index entry points.

Computer-assisted indexing techniques were developed to facilitate data input and to allow the generation of multiple access points from single input phrases by rearranging the order of the elements within multiword descriptors. These string manipulation procedures rely on data input format recognition software that minimizes keyboarding of duplicate data elements. The abbreviated record length in conjunction with the mechanized phrase manipulation software facilitated the execution of the

retrospective conversion of collection titles and reduces the time dedicated to data preparation, data entry and editing activities in the ongoing processing of new titles for cumulative supplements.

While the index is intended to be used primarily by reference librarians, it has proven to be easily approachable by patrons and other staff. For the sake of speed and convenience, the index uses a one-step approach; that is, an attempt is made to allow complete searches with one look-up in the index. The incorporation of multiple access routes for individual titles by providing entries under all significant elements of composite subjects and, in particular, the form designator(s) addresses the problem of lack of user perseverance in catalog searches. In one study of an academic library,[16] more than 50% of catalog users searched at only one entry point and then discontinued the search. Another study[17] found that multiple access greatly increases the chances of success in non-known-item searches.

Index Format and Record Structure

The index entries, illustrated in Appendix 1, consist of four fixed-length fields--lead term (designated as "subject"), title, date of publication and call number--arranged in a title-a-line 132 character KWOC style format. These elements are normally sufficient to locate a title on the shelf and can also serve as a link-key to the Card or Serials Catalog at which a patron can obtain more complete bibliographic information concerning a title. The short entry format is consistent with the fact that the primary focus of the index is to provide expanded subject access to reference materials. The KWOC style format offers index readability, and the single line entries proved necessary to keep the size of the index to a minimum. The display line presently allows 38 characters for subject descriptor and 57 characters for title. Approximately 5% of the subject descriptors and 9% of the titles are truncated at the output stage by the field length parameters. In the very few cases where index entries are not readily identifiable from the visible terms, abbreviations have been employed to overcome the problem.

The master records are variable length and contain title, descriptor terms, call numbers, date information and LC card number fields, all in varying length format. The abbreviated record format, of course, greatly expedited the retrospective conversion of reference titles into a machine-readable form. In order to link the short entry records to their MARC format analogs, the LC card number was included in those records for which the card number was readily available. This would allow augmenting the MARC record by adding descriptor strings to the 690 (local subject headings) field. The average length of the added descriptors field is approximately 115 characters. The compact data entry format greatly reduces duplication of data elements at the keyboarding stage and decreases the length of the input string. However, linkage between the reference collection index and MARC standard machine-readable bibliographic records is not envisioned at this time. The entire indexing project, including data preparation, data entry and production runs, is within the purview of the Reference Department.

Periodicals display varying information in the publication date field. Serials such as A & I services and book reviewing journals display the beginning holdings date and a continuation symbol, such as 1970+, in the publication date field. Items purchased on a yearly basis are assigned the publication date heading ANNUAL, while those publications which are frequently issued in new editions or are annual publications purchased by the Library on a staggered basis are given the heading UPDATED. The traditional edition statement is of little value for reference titles. Loose-leaf publications contain LOOSE-LEAF in the date of publication field. The use of these terms, in lieu of dates, minimizes the problems of updating the index for new editions and eliminates from the index any record-keeping function.

Indexing Vocabulary

As mentioned previously, the problem of access to reference materials was addressed in this project by providing (1) exhaustive subject analysis based on examination of the work, (2) title entries, (3) entry by significant words or phrases from the title, (4) author, compiler and/or editor, personal name entries on an infrequent basis, and (5) form or content designators as subdivisions under subjects

and as lead terms in descriptor strings.

The sample indexes revealed that, in most cases, individual reference titles can be categorized by various form or content designations, such as DICTIONARY, A & I SERVICE, BIBLIOGRAPHY, HANDBOOK, etc. A list of these identified form descriptors appears in Appendix 2. However, the form designator DIRECTORY was judged to be insufficient in specificity and not valuable as either a subdivision or lead term. Rather, one or more content descriptors (such as TRADE NAMES, BIOGRAPHICAL INFORMATION, etc.) or subject terms found in the work (such as COMPANIES, ASSOCIATIONS, FACULTY, etc.) are substituted. The form noun HANDBOOK is not used as a lead term and, while helpful as a subdivision, must be augmented with other content designators or subject terms. Consistent with recent trends in indexing, an uncontrolled vocabulary scheme using author's words taken directly from the work serves as the basis for the subject indexing system.[18] However, it was felt that retrieval was facilitated by combining subject or geographical terms with their appropriate form noun(s) to produce a precoordinated descriptor string.

The prescribed method of constructing subject descriptors is to identify the significant subject or geographical descriptors and combine them with the title's form/content indicator(s). These descriptors are in a passive order, natural language string, e.g., POLITICAL SCIENCE DICTIONARY or SIERRA LEONE STATISTICAL INFORMATION. An attempt was made to avoid the pitfalls of LCSH's multiplicity of punctuation symbols; entries in natural language contain no punctuation symbols, others contain hyphens at the point(s) where natural language order breaks, e.g., DICTIONARY--GERMAN-ENGLISH PHYSICS, DICTIONARY--PHYSICS--GERMAN-ENGLISH. Experience revealed that access by form/content terms was useful and outweighed any disadvantages of added size to the printed index. These form/content noun lead term entries are produced by phrase manipulation and will be discussed in the data entry section.

A problem of all uncontrolled vocabulary indexes is their scatter of variant forms (e.g., the terms Plan, Planned, Planning) and synonyms (e.g., Firms, Businesses, Business Enterprises, Companies, Corporations, etc.). The problem

of scatter due to indexer inconsistency is especially acute in applications such as this one in which a number of indexers are working separately without a thesaurus. For the retrospective conversion, a term normalization procedure operating on both single words and multiword phrases was introduced to address these problems. This procedure changes variant forms and synonyms into a single preferred term at the output stage. Additionally, syndetic structure routines tied to the normalization programs produce both "see" and "see-also" cross references and connect variant forms and related terms. The amount of normalization needed proved to be extensive and covered both subject terms and form/content nouns. This consumed a large amount of time in human editing of the raw output in order to build the authority file of terms coupled to their preferred forms. For the ongoing coding and entry of new titles, the index itself serves as a thesaurus of reference titles subject terminology, and the normalization routines are less extensively employed.

Another shortcoming of natural language indexing vocabularies is the lack of a hierarchial structure between descriptors (such as the BT, or broader term, and NT, or narrower term, designations in thesauri). In this index, the problem is overcome by abandoning the principle of specific entry in favor of exhaustive indexing at both the generic and specific levels. This is consistent with the policy of producing a one-step index. For example, a cell science dictionary is assigned both CELL SCIENCE and BIOLOGY as subject terms. A directory of midwestern manufacturers receives MIDWESTERN, WISCONSIN, ILLINOIS, MINNESOTA, etc., as geographic name headings. The term normalization procedures are used in many cases to accomplish this co-assignment of entry terms. In addition, the capability of providing cross references between terms helps to alleviate the impact of any subjective decisions about index term relationships.

Data Entry Software

To expedite the retrospective conversion of titles and to minimize coding errors, keypunching was done from Reference shelf list cards using data elements already present (title, call number, date of publication) along with strings and title word tags added, in pencil, to the cards by indexers. Appendix 3 shows several sample shelf list cards.

The data entry software admits two methods of indicating desired access points during input. The vast majority of lead terms are generated from data entered in the subject descriptors field. This variable length field can presently contain up to 2,000 characters, in the form of either single words or multiword strings. These descriptor strings are stream-coded and separated by semicolons.

The phrase manipulation software rearranges the elements of multiword strings in order to produce additional lead terms for the alphabetical index. The multiple entry points retain the whole of the compound subject within each of the lead terms. The rearrangement follows a prescribed manner and is dependent upon the positioning by an indexer of phrase manipulation symbols at desired points within the input string.

Keen[19-20] has summarized the multiple entry generation schemes, and this report will use Keen's definitions. The phrase manipulation procedures used in this project incorporate several of the mechanized indexing routines reported in the literature.

Three optional phrase manipulation symbols are available to the indexer. The symbol "#" produces a cycling operation in which the phrase elements following the "#" are brought to the lead position; a dash is then inserted and the elements preceding the "#" are added. For example, an input string of the form A#BC will yield ABC and BC--A as entries. Any number of manipulation symbols may be coded. The string A#B#C first yields ABC, BC--A and C--AB. In addition, any string with two or more "#", such as in the above example, will produce two additional articulated string entries generated by (1) combining the elements preceding the first "#" with those following the final "#" and subdividing with the remainder of the string, and (2) bringing the string elements following the final "#" to the lead, inserting the elements between the first and last "#" symbols, and following with the elements preceding the initial "#" symbol, with the three sections separated by hyphens. Thus, in the case of A#B#C, the two additional entries AC--B and C--B--A are generated by articulation. None of the phrase manipulation symbols will appear at the output stage.

As an example, the input string

GERMAN-ENGLISH#PHYSICS#DICTIONARY

generates the five lead terms:

GERMAN-ENGLISH PHYSICS DICTIONARY

PHYSICS DICTIONARY--GERMAN-ENGLISH

DICTIONARY--GERMAN-ENGLISH PHYSICS

GERMAN-ENGLISH DICTIONARY--PHYSICS

DICTIONARY--PHYSICS--GERMAN-ENGLISH.

The use of the manipulation symbol "+" causes a similar cycling operation, but a hyphen is inserted in the original string in place of the space. For example, the string

WOMEN+LITERATURE

produces the two lead terms:

WOMEN--LITERATURE

LITERATURE--WOMEN.

This symbol is used in the rare instances when the original input string cannot be constructed in a natural language, passive order phrase.

The cycling operation rearranges the order of the input string elements and brings significant elements of the string to the lead position, while preserving the contextual relationships of the term elements. The purpose of the articulation operation is to collocate those string element pairs that have become separated because of the length and form of the input string. These term pair coordinations are desirable in instances where other lead terms, either because of fewer constituent elements or a different arrangement, do display the particular pair combination. Without the articulation headings, an effective search would require the scanning of the entire body of entries under the first significant element in the lead term in order to insure finding all the combinations of the two elements. For example, the input string PHYSICS#DICTIONARY generates the lead term DICTIONARY--PHYSICS. In the previous example of GERMAN-ENGLISH#PHYSICS#DICTIONARY, the cycling

operation produced the lead term DICTIONARY--GERMAN-ENGLISH PHYSICS in which the elements DICTIONARY and PHYSICS are not directly coordinated. If the articulation operation were not performed, a search for all physics dictionaries approached via the form noun DICTIONARY would require reading all the lead terms under DICTIONARY to be certain of satisfying the search. However, the articulation operation produces DICTIONARY--PHYSICS--GERMAN-ENGLISH, and this term pair coordination collocates the two titles. Similarly, the articulation-produced lead entry GERMAN-ENGLISH DICTIONARY--PHYSICS insures coordination of all German-English dictionaries under the element GERMAN-ENGLISH.

It should be noted that for $n \geq 2$, a string containing N "#" symbols will generate N+1 lead terms by cycling, but only 2 additional leads by articulation for a total of N+3 entries. This serves to keep the number of index entries to a manageable number. For the same reason, it was decided that because of the passive order input string presentation, lead terms of the form B--A--C (corresponding to the input string A#B#C) were normally not valuable. For situations where non-standard types of coordinations are desired a truncation routine has been devised. This routine requires that a pair of brackets be placed around a subset of the entry input string, e.g., A#B[#C]. After the complete string is processed intact--that is, as if the brackets were not present--the elements and symbols within the brackets are eliminated, and the string is reprocessed. Thus, in the example above, the 5 cycling and articulation entries would be produced and, in addition, the 2 entries AB and B--A. This component truncation routine is most commonly used in conjunction with countries or generic--specific terms. For example, the string [CANADA]CHEMISTRY#FACULTY produces the lead terms

CANADA CHEMISTRY-FACULTY
FACULTY--CANADA CHEMISTRY
CHEMISTRY FACULTY
FACULTY--CHEMISTRY

and might be used for a directory of U. S. and Canadian university chemists. Another example of the use of truncation is in the subject term [SOLAR] ENERGY to generate both SOLAR ENERGY and ENERGY.

The second method for placing desired access points in the input stream is by tagging significant elements in the title field for removal and inclusion as subject heading terms. The software, after ignoring initial articles, also includes the full title as a lead term entry. For example, the coded title

WHO'S WHO IN#AMERICAN#ART

produces the lead terms

WHO'S WHO IN AMERICAN ART
AMERICAN ART--WHO'S WHO IN
ART--WHO'S WHO IN AMERICAN
ART--AMERICAN--WHO'S WHO IN
WHO'S WHO IN ART--AMERICAN.

The data preparation procedures are designed to minimize keyboarding strokes (and, hence, the incidence of keyboarding error) by greatly reducing duplication of entry string elements. This is possible because examination of sample indexes revealed a frequent word order pattern among groups of subject lead terms assigned to an individual title. A combinatorial based data input format was designed and the appropriate format recognition software written to accept input data and convert it to a form amenable to processing by the phrase manipulation algorithms. The result is a data input format that is much more compact than a traditional linear input format which would require a large amount of duplicate coding of phrase elements.

The data input format is predicated on the fact that a large number of the desired subject descriptors for reference titles are of the form subject/geographical identifier plus form noun(s). When applicable, two partitioned sets of descriptor elements are entered in the subject descriptors field in a data input string of the type $(A_1; A_2; \ldots; A_p)*(B_1; B_2; \ldots; B_q)$, where the A_i and B_i are single words or multiword phrases and "*" is a phrase manipulation symbol. Elements within

a set are separated by semicolons. The partitioned sets input string is then expanded by the format recognition routine in a distributive manner to yield the pxq multiword strings A_1*B_1, A_1*B_2, . . ., A_1*B_q, A_2*B_1, A_2*B_2, . . ., A_2*B_q, . . ., A_p*B_1, A_p*B_2, . . ., A_p*B_q. These Discrete strings are then passed individually to the phrase manipulation software to generate lead terms for index entries. For example, the input string (PHYSICS; MATHEMATICS)# (DICTIONARY; BIOGRAPHICAL INFORMATION) sends the input strings

PHYSICS#DICTIONARY

PHYSICS#BIOGRAPHICAL INFORMATION

MATHEMATICS#DICTIONARY

MATHEMATICS#BIOGRAPHICAL INFORMATION

to the phrase manipulation routine to produce 8 lead terms for the index. Individual elements within the descriptor sets may contain embedded articulation symbols. These are passed intact to the phrase manipulation software.

The third manipulation symbol "-" is most often used in conjunction with the partitioned sets format. This operator's function is to signal the program that no manipulation should be performed. For example, the string (UNITED NATIONS; INTERNATIONAL ORGANIZATIONS)-(OFFICIALS) produces the 2 lead terms

UNITED NATIONS OFFICIALS

INTERNATIONAL ORGANIZATIONS OFFICIALS

The symbol "-" can be used to shorten input strings under certain conditions. For example, the input form (GERMAN-; FRENCH-; RUSSIAN-)-(ENGLISH# PHYSICS#DICTIONARY) generates the same lead terms as (GERMAN-ENGLISH; FRENCH-ENGLISH; RUSSIAN-ENGLISH)#(PHYSICS#DICTIONARY) but does not require the repeated entry of the term ENGLISH.

Several specialized software routines have been introduced to further expedite multiple entry generation and provide a more compact data input format. The symbol "≠", when placed before a term element, indicates to the phrase manipulation routine that the term should not be moved to the lead position. For example, the string (CIVIL ENGINEERING)* (TABLES;≠HANDBOOK) ultimately produces three lead terms:

CIVIL ENGINEERING TABLES
CIVIL ENGINEERING HANDBOOK
TABLES--CIVIL ENGINEERING.

The symbol "≠" is effectively used in front of individual term elements in situations where some, but not all, of the elements in either descriptor set merit lead element status.

The term normalization routines are employed to streamline data input in two ways. Heavily used terms, such as the form nouns, can be normalized from abbreviated or mnemonic forms. For example, the descriptor authority file contains the word DY as a variant form of the term DICTIONARY. Thus, from an earlier example, we can write the input string (PHYSICS; MATHEMATICS)#(DY;BI), where BI is normalized to BIOGRAPHICAL INFORMATION. The normalization software is also utilized to replace an individual term with itself and another coupled term, usually a term hierarchially related. For example, the term RAILROADS cues the normalization program to substitute RAILROADS; TRANSPORTATION in its stead. In this way, the broader heading will also appear as a separate lead term element. However, the use of the term TRANSPORTATION will not automatically produce the above substitution.

The number of lead terms produced by a partitioned sets string can be determined, but the formula depends on the particular manipulation symbol included in the string. In instances where the symbol "#" separates the two sets, the number N of lead terms produced can be determined from the formula $N=2xpq+S_B xp+S_A xq+R$, where p is, as before, the number of elements in the first set $\{A_i\}$, q the number of elements in the second set $\{B_i\}$, S_B the total number of "#" symbols within the elements of the set $\{B_i\}$, S_A the total number of "#" symbols within the elements of the set $\{A_i\}$, and R is the total number of lead terms produced by articulation. The number R can only be determined by inspection of the partitioned sets. However, in situations where each element of $\{B_i\}$ contains at least one

"#" or each element of $\{A_i\}$ contains at least one embedded "#", R=2xp. If the symbol "-" separates the two partitioned sets, the formula $N=pxq + S_B xp + S_A xq + R$ holds.

The software for this project was written in PL/I F level and optimizer compiler. The main procedure contains four modules: (1) a format recognition routine to translate input strings for processing, (2) a word normalization routine, (3) a cross-reference production procedure, and (4) the phrase manipulation and entry production module. A copy of the source program is included as Appendix 4. The data entry procedures and/or phrase manipulation algorithm could be used in any indexing application desiring a combinatorial-based data entry format or requiring a multiple entry generation procedure capable of allowing controlled combinations of input string elements.

Access Points

The need for certain types of access points was identified primarily from anecdotal evidence compiled by reference staff. Likewise, much of the terminology used in the subject descriptors was derived from experiences using LCSH and other subject heading schemes. In order to better illustrate the raison d' être for this index, selected titles that typify these types of access points will be presented and any differences with LCSH examined.

Title access allows easy identification of the call number in those instances where known-item title searches are performed, usually by reference librarians. Access by significant title words is helpful when the complete title is unknown or an incorrect citation has been provided. For example, a search for a "Who's Who in Music" under that title will locate the heading WHO'S WHO IN MUSIC--INTERNATIONAL, a retrieval key for the title International Who's Who in Music. Likewise, the reference collection index contains an entry BOOK OF LISTS for the title whose main entry is People's Almanac Presents the Book of Lists.

Form/content descriptors prove useful in situations where there may not be an exact match for a search request. For example, a search for a plant science dictionary yields nothing under that heading but searching under the subdivisions at the form descriptor DICTIONARY yields DICTIONARY--FORESTRY and DICTIONARY--BOTANY with associated relevant titles. This type of form descriptor access is not possible in LCSH.

Personal or corporate name entries are indicated for works closely associated with a person or body such as HAZLITT for the Bibliography of Early English Literature 1867-1903 or CHICAGO--UNIVERSITY OF and UNIVERSITY OF CHICAGO for the title A Manual of Style which is associated with that institution.

The real value of the index lies in its exhaustive subject access to reference titles. Specific features of titles are brought out; e.g., ENGINEERING AWARDS for Who's Who in Engineering, AGRICULTURAL ENGINEERING STANDARDS in Agricultural Engineers Yearbook, and DANCE UNIVERSITIES from The National Directory for the Performing Arts--Education. The lack of indexing exhaustivity is a major problem in LCSH. To illustrate, the title Dictionary of Scandinavian Biography (LCCN 73-189270) was assigned the two subject headings SCANDINAVIA--BIOGRAPHY--DICTIONARIES and FINLAND--BIOGRAPHY--DICTIONARIES from LCSH. However, the title also includes biographical information for Norway, Denmark, Sweden, and Iceland. The reference collection index contains the twelve retrieval points generated from the data entry string (SCANDINAVIA; NORWAY; FINLAND; DENMARK; ICELAND; SWEDEN)#(BIOGRAPHICAL INFORMATION) and two additional access points from the additional input phrase SCANDINAVIA COOPERATION#TREATY for a total of fourteen subject entries. Examination of some of the traditional reference tools, such as Europa Yearbook (ISSN 71-2302) reveals the wide discrepancies in coverage. For the reference collection index, this title has been assigned the following string in the subject descriptors field:

> (INTERNATIONAL ORGANIZATIONS; ECONOMIC#COMMISSIONS; UNITED NATIONS)-(OFFICIALS); (ALBANIA; ANDORRA; AUSTRIA; etc., the European countries through Yugoslavia; EUROPE)# ([ECONOMIC#]STATISTICAL INFORMATION; ALMANAC; AGRICULTURE STATISTICAL INFORMATION; CONSTITUTIONS; GOVERNMENT OFFICIALS; EMBASSIES; NEWSPAPERS; PERIODICALS; PUBLISHERS; BANKS; ASSOCIATIONS; LABOR UNIONS; UNIVERSITIES).

From this some 300 subject entries are produced. Contrast this with the two LCSH assigned to the title:

> EUROPE--POLITICS--YEARBOOK
> EUROPEAN FEDERATION--YEARBOOKS.

The scatter of titles under synonyms and related terms is another major problem area in using LCSH to retrieve reference titles. For example, National Faculty Directory (ISSN 77-4472) has been given the heading COLLEGE TEACHERS--UNITED STATES--DIRECTORIES; Directory of American Scholars (ISSN 70-5101) has the heading SCHOLARS, AMERICAN--DIRECTORIES; American Men and Women of Science (ISSN 192-8576) the heading SCIENTISTS--UNITED STATES; AIBS Directory of Bioscience Departments and Faculty (LCCN 75-33761) has been assigned BIOLOGY--STUDY AND TEACHING--U.S.--DIRECTORIES; and American Men and Women of Science. Biology (ISSN 146-0048), in contrast to the previous title, was assigned BIOLOGISTS--U.S.--BIOGRAPHY from LCSH. All of these titles will be found under FACULTY, in addition to other access points, in the reference collection index.

As another example, consider several comprehensive business directories. Consumer's Register of American Business (ISSN 98-7344) bears the heading BUSINESS ENTERPRISES--U.S.--DIRECTORIES from LCSH. However, Moody's Industrial Manual (ISSN 545-0217) has been assigned CORPORATIONS--U.S.--DIRECTORIES; the Standard and Poor's Register of Corporations, Directors, and Executives (LCCN 28-7849) has been given only DIRECTORS OF CORPORATIONS--U.S.--DIRECTORIES and the identical heading with CANADA substituted for U.S.; and the Dun and Bradstreet Million Dollar Directory (LCCN 59-3033) bears the LC subject heading U.S.--INDUSTRIES--DIRECTORIES. These titles appear under the term COMPANIES in the reference collection index.

LCSH's use of outmoded terminology also complicates retrieval. Energy Technology Handbook (LCCN 76-17653) has the headings POWER(MECHANICS)--HANDBOOKS and POWER RESOURCES--HANDBOOKS from LCSH as opposed to the fifty headings produced by the input string (ENERGY; COAL; PETROLEUM; NUCLEAR ENERGY; SOLAR ENERGY; GEOTHERMAL ENERGY; GASOLINE; HYDROPOWER; POWER ENGINEERING; INDUSTRIAL ENGINEERING)#(PHYSICAL CONSTANTS; ≠HANDBOOK; DICTIONARY) for the reference collection index.

These examples clearly demonstrate the significant problems of effectively retrieving reference titles by subject using LCSH. These problems have been addressed in this project by providing expanded subject analysis in a printed format display. An average of approximately 11 access points has been assigned per title;

for titles processed during the last year the average has been over 16 access points, compared to an average 1.8 subject headings from subject cataloging done at Library of Congress.[21]

Conclusion

This project demonstrated the feasibility of a short entry printed index emphasizing expanded subject access to academic library reference collection materials. In terms of bibliographic role, this index is designed to complement, not replace, other library catalogs. The index provides access points not present in the other catalogs, indicates special locations within the reference collection (e.g., index tables), covers bibliographies (e.g., Reader's Advisory Service) and series (Midwest Plan Service publications) not analyzed in the other catalogs, and identifies the A & I services that can be searched through an on-line vendor. While the index is most heavily used by reference librarians, it has proven to be easily approachable by patrons and other staff. It has been a valuable tool for training new reference staff and assisting paraprofessionals. The index is produced in both hardcopy and microfiche, and presently totals more than 1,000 pages and 70,000 entries. The index guide's portability allows it to be placed in branch library reading rooms and other locations in the library system. A separate index covering only the A & I services is produced as a spin-off. As for its adaptability by other institutions, while it may be helpful in reference service, the differences in holdings, call numbers, and location designations between Iowa State University and other academic libraries would limit its usefulness. However, the procedures utilized in this project could be transferred to other settings and successfully applied to either a reference collection or any other component of a library collection.

A copy of the most recent complete printed index is included as Appendix 5.

Bibliography

1. Cherie Weil, "Augomatic Retrieval of Biographical Reference Tools," Journal of Library Automation 1:239-49 (December 1968).

2. N. La Var Harline, A Computerized Reference Retrieval System ED 102 974 (Brigham Young University. Graduate Department of Library and Information Sciences, 1974).

3. Ching-Chih Chen and E. Robert Kingham, "Subject Reference Lists Produced by Computer," Journal of Library Automation 1:178-97 (September 1968).

4. J. C. Meredith, "Machine-Assisted Approach to General Reference Materials," Journal of the American Society for Information Science 22:171-86 (May-June 1971).

5. James Krikelas, "Catalog Use Studies and their Implications," Advances in Librarianship 3:195-220 (1972).

6. Hans Wellisch, "Subject Retrieval in the Seventies," in Hans Wellisch and Thomas D. Wilson, ed., Subject Retrieval in the Seventies (Westport, Conn.: Greenwood Publishing Co., 1972), pp. 5-15.

7. Sanford Berman, "The Cataloging Shtik," Library Journal 102:1251-1253 (June 1, 1977).

8. Marcia J. Bates, "Factors Affecting Subject Catalog Search Success," Journal of the American Society for Information Science 28:161-169 (May 1977).

9. Lois Mai Chan, "Alphabetical Arrangement and Subject Collocation in Library of Congress Subject Headings," Library Resources and Technical Services 21:156-169 (Spring 1977).

10. William H. Mischo, "An Augmented Keyword Index to Abstracting and Indexing Services," Journal of Academic Librarianship 3:141-145 (July 1977).

11. William H. Mischo, "Library of Congress Subject Headings: A Review of the Problems, and Prospects for Improved Subject Access," (Manuscript out for review).

12. Kieth C. Wright, "Trends in Modern Subject Analysis with Reference to Text Derivative Indexing and Abstracting Methods; The State of the Art," Information Part 2 1:6 (September-October 1972).

13. Pauline Atherton, Books are for Use: Final Report of the Subject Access Project to the Council on Library Resources (Syracuse, N.Y.; Syracuse University School of Information Studies, 1978), ED 156 181.

14. S. Michael Malinconico, "The Display Medium," Library Journal 101:2144-49 (1976).

15. R. P. Palmer, Computerizing the Card Catalog in the University Library: A Survey of User Requirements (Littleton, Colo.: Libraries Unlimited, Inc., 1972).

16. R. Tagliacozzo and M. Kochen, "Information-Seeking Behavior of Catalog Users," Information Storage and Retrieval 6:375 (December 1970).

17. D. R. Swanson, "Requirements Study for Future Catalogs," Library Quarterly 42:302-315 (1972).

18. F. W. Lancaster, "Vocabulary Control and Information Retrieval Systems," Advances in Librarianship 8:2-40 (1977).

19. E. Michael Keen, On the Performance of Nine Printed Subject Index Entry Types (Aberstwyth, Wales: College of Librarianship Wales, 1978).

20. E. Michael Keen, "On the Generation and Searching of Entries in Printed Subject Indexes," Journal of Documentation 33:15-45 (March 1977).

21. Martha E. Williams and Gordon J. Shefner, "Data Element Statistics for the MARC II Data Base," Journal of Library Automation 9:89-100 (June 1976).

REFERENCE COLLECTION INDEX SUBJECT HEADING	TITLE	DATE, CALL NUMBER
POLLUTANTS-CHEMICAL	HANDGEN INDEX	UPDATED SB951 .P23
POLLUTION	INDUSTRIAL POLLUTION CONTROL HANDBOOK	1971 TD897 .I42
POLLUTION ABSTRACTS	**POLLUTION ABSTRACTS	TD180 A1P6 V.1, 1970+
POLLUTION ASSOCIATIONS	POLLUTION RESEARCH INDEX	1975 TD178.5 .P64
POLLUTION ASSOCIATIONS	U.S. LIBRARY OF CONGRESS DIRECTORY OF INFORMATION SERVICE	1966 TD211 .UN3D
POLLUTION ASSOCIATIONS	SOURCEBOOK ON THE ENVIRONMENT	1978 GF41 .A11 S69
POLLUTION ASSOCIATIONS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION ASSOCIATIONS	ENVIRONMENTAL LAW	1977 KF3775 .A1 S35
POLLUTION ASSOCIATIONS-LAND	DIRECTORY OF NATIONAL ORGANIZATIONS CONCERNED WITH LAND	1974 TD173.5 .D5X 1970
POLLUTION BIBLIOGRAPHY	DICTIONARY CATALOG OF THE WATER RESOURCES CENTER ARCHIVES	1970 GB661 .A11C3X
POLLUTION BIBLIOGRAPHY	ENVIRONMENTAL LAW	1977 KF3775 .A1 S35
POLLUTION BIBLIOGRAPHY	SOURCEBOOK ON THE ENVIRONMENT	1978 GF41 .A11 S69
POLLUTION BIBLIOGRAPHY	WASTE WATER MANAGEMENT	1976 TD741 .A11 T45
POLLUTION BIOLOGY HANDBOOK	HANDBOOK ON ENVIRONMENTAL DATA ON ORGANIC CHEMICALS	1977 TD196 .O73 V47
POLLUTION COMPANIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION CONSULTANTS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION CONTROL GUIDE	POLLUTION CONTROL GUIDE	LOOSE LEAF KF3775 .A6 C65
POLLUTION CONTROL-CHEMICALS-HAZARDOUS	HANDLING GUIDE FOR POTENTIALLY HAZARDOUS MATERIALS	1975 T55.3 .H3 H35X
POLLUTION CONTROL-HAZARDOUS CHEMICALS	HANDLING GUIDE FOR POTENTIALLY HAZARDOUS MATERIALS	1975 T55.3 .H3 H35X
POLLUTION ENCYCLOPEDIA	THE COMPLETE ECOLOGY FACT BOOK	1972 TD174 .N6
POLLUTION ENGINEERING	POLLUTION ENGINEERING PRACTICE HANDBOOK	1975 TD151 .P64
POLLUTION FORMULAS	HANDBOOK ON ENVIRONMENTAL DATA ON ORGANIC CHEMICALS	1977 TD196 .O73 V47
POLLUTION GOVERNMENT AGENCIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION GOVERNMENT AGENCIES	POLLUTION RESEARCH INDEX	1975 TD178.5 .P64
POLLUTION GOVERNMENT AGENCIES-AIR	DIRECTORY OF GOVERNMENTAL AIR POLLUTION AGENCIES	UPDATED TD882 .D5
POLLUTION GOVERNMENT AGENCIES-LAND	DIRECTORY OF NATIONAL ORGANIZATIONS CONCERNED WITH LAND	1974 TD173.5 .D5X 1970
POLLUTION HANDBOOK	ENVIRONMENTAL SOURCES AND EMISSIONS HANDBOOK	1975 TD897.5 .S58
POLLUTION INDEX	DICTIONARY CATALOG OF THE WATER RESOURCES CENTER ARCHIVES	1970 GB661 .A11C3X
POLLUTION INDEX/ABSTRACT SERVICE	**POLLUTION ABSTRACTS	TD180 A1P6 V.1, 1970+
POLLUTION INDEX/ABSTRACT SERVICE	SELECTED WATER RESOURCES ABSTRACTS	GB651 S4 V.2, 1969+
POLLUTION INDEX/ABSTRACT SERVICE	ANNOTATED BIBLIOGRAPHY ON HYDROLOGY AND SEDIMENTATION	GB652 A835A 1941-1968
POLLUTION INDEX/ABSTRACT SERVICE	**AIR POLLUTION ABSTRACTS	TD883 A11 A38 1970-1976
POLLUTION INDEX/ABSTRACT SERVICE	**AQUATIC SCIENCES AND FISHERIES ABSTRACTS (BIOLOGICAL SCI	QH90 A1 A633 V.4, 1974+
POLLUTION INDEX/ABSTRACT SERVICE	APPLIED ECOLOGY ABSTRACTS	QH540 A6 V.1, 1975+
POLLUTION INDEX/ABSTRACT SERVICE	BIBLIOGRAPHY OF NOISE	TD892 A11 F58 1965+
POLLUTION INDEX/ABSTRACT SERVICE	ABSTRACTS ON HEALTH EFFECTS OF ENVIRONMENTAL POLLUTANTS	RA565 A1 A2 V.1,1972+
POLLUTION INDEX/ABSTRACT SERVICE	WATER POLLUTION ABSTRACTS	TD3 G798S V.3,1930-1973
POLLUTION INDEX/ABSTRACT SERVICE	WATER RESOURCES ABSTRACTS	GB651 W29 V.1,1968+
POLLUTION INDEX/ABSTRACT SERVICE	LAND USE PLANNING ABSTRACTS	HD171 A1 L355 1970+
POLLUTION INDEX/ABSTRACT SERVICE	INTERNATIONAL PETROLEUM ABSTRACTS (CON. INST. OF PETROLEU	TN860 IN7JAL V.39, 1953+
POLLUTION INDEX/ABSTRACT SERVICE	ENERGY INDEX	HD9502 US E525 1970+
POLLUTION INDEX/ABSTRACT SERVICE	**ENVIRONMENT INDEX	TD172 E52 V.1 1971+
POLLUTION INDEX/ABSTRACT SERVICE	EUTROPHICATION: SUMMARY OF CURRENT LITERATURE	QH96.8 E9W58 V.1-48,1969-197
POLLUTION INDEX/ABSTRACT SERVICE	CURRENT GEOGRAPHICAL PUBLICATIONS	G5 AM35C V.13, 1950+
POLLUTION INDEX/ABSTRACT SERVICE	**EXCERPTA MEDICA, SECTION 46: ENVIRONMENTAL HEALTH AND POL	SEARCHED BY COMPUTER
POLLUTION INFORMATION SERVICES	U.S. LIBRARY OF CONGRESS DIRECTORY OF INFORMATION SERVICE	1966 TD211 .UN3D
POLLUTION INSTITUTES-RESEARCH	POLLUTION RESEARCH INDEX	1975 TD178.5 .P64
POLLUTION INSTITUTES-RESEARCH	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION INTEREST GROUPS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION LABORATORIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION LAW	SOURCEBOOK ON THE ENVIRONMENT	1978 GF41 .A11 S69
POLLUTION LAW	POLLUTION CONTROL GUIDE	LOOSE LEAF KF3775 .A6 C65
POLLUTION LAW	ENVIRONMENTAL ENGINEERS' HANDBOOK	1974 TD145 .E57
POLLUTION LAW FIRMS	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65
POLLUTION LIBRARIES	WORLD ENVIRONMENTAL DIRECTORY	UPDATED TD12 .W65

** THIS ABSTRACTING OR INDEXING SERVICE CAN BE SEARCHED BY COMPUTER. INQUIRE AT THE REFERENCE DESK.

Appendix 2

Form and Content Designators

Abstracting and Indexing Service (A & I Service)

Almanac

Atlas

Bibliography

Bio-Bibliography

Biographical Information

Book Catalog

Book Reviews

Catalog

Collections (used for book or material collections)

Concordance

Dictionary

Directory

Dissertations

Encyclopedia

Gazetteer

Handbook

Historical Atlas

Literature Guide

Maps

Occupational Information

Periodicals (list)

Physical Constants

Reference Works (bibliography)

Statistical Information

Style Manual

Tables

Thesaurus

Theses

Trade Names

Travel Information

Union List

Appendix 3

A4 RSL

TJ163.8
E54 (Energy technology handbook) / prepared by 142 specialists ; Douglas M. Considine, editor-in-chief. New York : McGraw-Hill, c1977.
1884 p. in various pagings : ill. ; 24 cm.
Includes bibliographical references and index.

(Coal; petroleum; gas; chemical engineering; nuclear science; solar energy; energy; geothermal energy; hydropower; power engineering; industrial engineering) #(Handbook; Dictionary; Physical Constants)

1. Power (Mechanics)--Handbooks, manuals, etc. 2. Power resources--Handbooks, manuals, etc. I. Considine, Douglas Maxwell.

IaAS 21 NOV 77 2238305 IWARsl 76-17653

LB15
H66 OK 1968 RSL

Hopke, William E
Dictionary of personnel and guidance terms, including professional agencies and associations, edited by William E. Hopke. [1st ed.] Chicago, J. G. Ferguson Pub. Co. [1968]
xix, 464 p. 25 cm.
Bibliography: p. 399-421.
1. Education—Terminology. I. Title. II. Title: Personnel and guidance terms.
LB15.H66 371.4'03 68-57491

(Personnel; guidance; education) #(Associations; state agencies; dictionary)

CT1243
D5x (Biog-E)

Dictionary of (Scandinavian biography) / Hon. general editor: Ernest Kay. With a memoir on the work of the Nordic Council by G. F. D. Dawson, and with the full text of the treaty of cooperation between Denmark, Finland, Iceland, Norway and Sweden. London. Melrose Press [1972]
xxxv, 467 p. 25 cm. £10.50 B***

(Scandinavia; Finland; Denmark; Iceland; Norway; Sweden) #(biographical Information); Scandinavia cooperation #Treaties

1. Scandinavia — Biography — Dictionaries. 2. Finland — Biography—Dictionaries. I. Kay, Ernest, ed. II. Denmark. Treaties, etc., 1947-1972. (Frederik IX) Treaty of cooperation between Denmark, Finland, Iceland, Norway and Sweden, 1971.

CT1243.D53 920'.048 [B] 73-189270
ISBN 0-900332-30-4 MARC
Library of Congress 72-4

```
KWOC: PROC OPTIONS(MAIN);
    DCL NCE BIT(1) INIT('0'B);DCL MCE BIT(1) INIT('0'B);
    DCL LLL FIXED;
     DCL CALLN CHAR(22);
    DCL W2 CHAR(80) VAR;
    DCL NORML(150) CHAR(55),II FIXED BINARY,STI CHAR(461) VAR,(MID,
     HI,LO) FIXED BINARY, SWIT BIT(1);
     DCL NEPTUN FILE RECORD OUTPUT;
     DCL W3 CHAR(800) VAR,M1 FIXED,STORE1 CHAR(80) VAR,(STR1,ST32,
     STR52)  CHAR(500) VAR,STORE2 CHAR(80) VAR;
   DCL K4 FIXED, DATE CHAR(10);
       DCL XKA PIC '(5)Z9';
    DCL N FIXED BIN(15);
   DCL JET CHAR(57),TI CHAR(861) VAR,W CHAR(80) VAR;
  DCL KAL_EL CHAR(861) VAR,CHECK BIT(1),
     DYLAN LABEL,
     BOGIE CHAR(80),
      JS FIXED,JT FIXED,K1 FIXED,K2 FIXED,K3 FIXED,CW CHAR(132) VAR,
    WORD CHAR(132),KA FIXED,FILL CHAR(240) INIT((240)' '),(L,K,I)
     FIXED;
  DCL CTI CHAR(250) VAR;
     DCL (L3,L4) FIXED;
   DCL ONE CHAR(132) INIT('HERE FOLLOW THE BOGUS INPUT SET');
DCL 1 CHANG,
      2 FIRST CHAR(1) INIT(' '),
      2 PITH CHAR (132);
 DCL NEPTON FILE RECORD OUTPUT,SYSIN FILE RECORD INPUT;
     DCL SORTIN FILE RECORD OUTPUT;
      ON ENDFILE(SYSIN) GO TO NEILS;
      PITH=ONE;
      WRITE FILE(NEPTON) FROM(CHANG);
     PITH=FILL;
      N=0;
    II=0;
    I=1;
    READ FILE(SYSIN) INTO(BOGIE);
     DO WHILE(SUBSTR(BOGIE,1,1)='0');
     L=INDEX(BOGIE,'#');
     IF L=0 THEN DO;
        PITH=BOGIE;
         WRITE FILE(NEPTON) FROM(CHANG);
        END;
      ELSE DO;
     NORML(I)=SUBSTR(BOGIE,2);
     I=I+1;
     END;
     READ FILE(SYSIN) INTO(BOGIE);
     END;
     II=I-1;
      READ FILE(SYSIN) INTO(BOGIE);
    DO WHILE(SUBSTR(BOGIE,1,1)='7');
     WORD=(3)' '||SUBSTR(BOGIE,2);
     WRITE FILE(SORTIN) FROM(WORD);
      READ FILE(SYSIN) INTO(BOGIE);
        END;
FETCH: DO I=1 BY 1;
      CHECK='0'B;
      KAL_EL=' ';
KARMA:IF CHECK='1'B THEN DO;
 UFO: IF LENGTH(KAL_EL)>132 THEN DO;
```

```
        PITH=SUBSTR(KAL_EL,1,132);
        KAL_EL=SUBSTR(KAL_EL,133);
      DYLAN=UFO;
        END;
        ELSE DO;
        PITH=FILL;
        PITH=KAL_EL;
        DYLAN=BLUE;
        END;
    WRITE FILE(NEPTON) FROM(CHANG);
        GO TO DYLAN;
BLUE:KAL_EL=' ';
        READ FILE(SYSIN) INTO(BOGIE);
        KAL_EL=KAL_EL||BOGIE;
      IF SUBSTR(BOGIE,1,1)¬='%' THEN GO TO KARMA;
        CHECK='0'B;
       IF VERIFY('%/6*#=',KAL_EL)=0 THEN GO TO PLUTO;
       END;
      DO WHILE(VERIFY('%/6*#=',KAL_EL)¬=0);
        READ FILE(SYSIN) INTO(BOGIE);
        KAL_EL=KAL_EL||BOGIE;
      IF LENGTH(KAL_EL)=80 THEN DO;
            CHECK='1'B;
        GO TO KARMA;
        END;
       END;
PLUTO:    IF SUBSTR(KAL_EL,2,1)¬='%'|VERIFY('%',SUBSTR(KAL_EL,3))=0
      THEN DO;
      CHECK='1'B;
      GO TO KARMA;
       END;
      K1=INDEX(KAL_EL,'/');
       K2=INDEX(KAL_EL,'6');
         K3=INDEX(KAL_EL,'*');
      K4=INDEX(SUBSTR(KAL_EL,K2+1),'#');
           K=K3-K2-1;
      IF (K2-K1)<0 THEN DO;
            CHECK='1'B;
        GO TO KARMA;
        END;
      IF (K2-K1-1)=0 THEN TI='';
       ELSE TI=SUBSTR(KAL_EL,K1+1,K2-K1-1);
            IF INDEX(KAL_EL,'@')¬=0 THEN DO;
        IF  SUBSTR(KAL_EL,K2+K4-4,1)='1' THEN
          DATE=SUBSTR(KAL_EL,K2+K4-4,4);
         ELSE IF SUBSTR(KAL_EL,K2+K4-1,1)='+' THEN
                DATE=SUBSTR(KAL_EL,K2+K4-5,5);
    ELSE IF SUBSTR(KAL_EL,K2+K4-1,1)='L' THEN
                DATE=SUBSTR(KAL_EL,K2+K4-6,6);
       ELSE IF SUBSTR(KAL_EL,K2+K4-1,1)='D' THEN
              DATE=SUBSTR(KAL_EL,K2+K4-7,7);
       ELSE IF SUBSTR(KAL_EL,K2+K4-1,1)='F' THEN
          DATE=SUBSTR(KAL_EL,K2+K4-10,10);
           ELSE DATE=' ';
           END;
             ELSE DATE=SUBSTR(KAL_EL,K3+1,K2+K4-K3-1);
     L=INDEX(SUBSTR(KAL_EL,K2+1,45),'(');
     IF L¬=0 THEN CALLN='REF DESK '||SUBSTR(KAL_EL,K2+1,L-1);
       ELSE CALLN=SUBSTR(KAL_EL,K2+1,K3-K2-1);
       N=N+1;
```

```
IF II¬=0 THEN DO;
IF TI¬='' THEN DO;
STI=TRANSLATE(TI,'        ',';-@<>$%');
L=1;
K1=0;
   SBIT='0'B;
DO WHILE(L¬=0);
 L=INDEX(STI,' ');
IF L=1 THEN DO;
 IF LENGTH(STI)>1 THEN DO;
    IF SUBSTR(STI,2,1)¬=' ' THEN DO;
         K1=K1+1;
         END;
       ELSE L=0;
   END;
 ELSE L=0;
  END;
 ELSE DO;
 IF L=0 THEN W=STI;
   ELSE W=SUBSTR(STI,1,L-1);
    LO=1;
    HI=II ;
     CHECK='0'B;
     DO WHILE(CHECK='0'B & (HI-LO)>0);
       MID=(LO+HI)/2.0;
   JT=INDEX(NORML(MID),';');
 W2=SUBSTR(NORML(MID),1,JT-1);
IF W<W2 THEN
            HI=MID-1;
 ELSE IF W>W2 THEN LO=MID+1;
ELSE CHECK='1'B;
 END;
IF (HI-LO)<=0 THEN DO;
       JT=INDEX(NORML(HI),';');
 W2=SUBSTR(NORML(HI),1,JT-1);
 IF W=W2 THEN DO;
      JS=INDEX(NORML(HI),'@');
     IF  (JS-JT)=1 THEN W2='';
 ELSE W2=SUBSTR(NORML(HI),JT+1,JS-JT-1);
      SBIT='1'B;
      END;
   END;
  ELSE DO;
       JS=INDEX(NORML(MID),'@');
       JT=INDEX(NORML(MID),';');
     IF  (JS-JT)=1 THEN W2='';
       ELSE W2=SUBSTR(NORML(MID),JT+1,JS-JT-1);
        SBIT='1'B;
 END;
     DO WHILE(SBIT='1'B);
 IF W2='' THEN DO;
   IF L¬=0 THEN DO;
 IF (SUBSTR(TI,K1,1)='$' | SUBSTR(TI,K1,1)='<'|
   SUBSTR(TI,K1,1)='>')
       THEN SUBSTR(TI,K1)=SUBSTR(TI,K1,1)||SUBSTR(TI,K1+L+1);
   ELSE SUBSTR(TI,K1)=SUBSTR(TI,K1+L);
 END;
  ELSE SUBSTR(TI,K1)=FILL;
  K1=K1-L;
 END;
```

```
      ELSE DO;
        IF L=0 THEN SUBSTR(TI,K1+1)=W2||FILL;
        ELSE SUBSTR(TI,K1+1)=W2||SUBSTR(TI,K1+L);
        K1=K1+(LENGTH(W2)-(L-1));
        END;
          SBIT='0'B;
          END;
          K1=K1+L;
        END;
         IF L¬=0 THEN STI=SUBSTR(STI,L+1);
         END;
      END;
      END;
       K1=INDEX(KAL_EL,'/');
      LLL=INDEX(TI,'STATISTICAL INFORMATION');
      DO WHILE(LLL¬=0);
     TI=SUBSTR(TI,1,LLL-1)||'STATS INFO'||SUBSTR(TI,LLL+23);
      LLL=INDEX(TI,'STATISTICAL INFORMATION');
      END;
      LLL=INDEX(TI,'BIOGRAPHICAL INFORMATION');
      DO WHILE(LLL¬=0);
      TI=SUBSTR(TI,1,LLL-1)||'BIOG INFO'||SUBSTR(TI,LLL+24);
      LLL=INDEX(TI,'BIOGRAPHICAL INFORMATION');
      END;
       JET=SUBSTR(KAL_EL,3,K1-3) ;
      JET=TRANSLATE(JET,'    ',',()#');
      CTI=SUBSTR(KAL_EL,3,K1-2);
       L3=INDEX(CTI,'(');
       DO WHILE(L3¬=0);
        IF L3=1 THEN JET=SUBSTR(JET,2);
      DO WHILE(L3¬=0);
      L4=INDEX(CTI,')');
      IF L4=0 THEN DO;
      CHECK='1'B;
      GO TO KARMA;
       END;
       TI=SUBSTR(CTI,L3+1,L4-L3-1)||':'||TI;
      CTI=SUBSTR(CTI,L4+1);
        L3=INDEX(CTI,'(');
        END;
      END;
       CW=JET||(1)' '||DATE||(1)' '||CALLN;
     PITH=CW;
       WRITE FILE(NEPTON) FROM(CHANG);
      IF SUBSTR(KAL_EL,3,1)¬='(' THEN DO;
         IF SUBSTR(KAL_EL,3,3)='THE' THEN DO;
         IF SUBSTR(KAL_EL,6,1)¬='(' THEN DO;
     IF SUBSTR(KAL_EL,6,1)=' ' THEN DO;
          W=SUBSTR(KAL_EL,7,K1-7);
      W=TRANSLATE(W,'    ',',()#');
           CALL DWRIT(W,CW);
          END;
          END;
          END;
      ELSE IF SUBSTR(KAL_EL,3,1)='A' THEN DO;
         IF SUBSTR(KAL_EL,4,1)¬='(' THEN DO;
         IF SUBSTR(KAL_EL,4,1)=' ' THEN DO;
         W=SUBSTR(KAL_EL,8,K1-5);
      W=TRANSLATE(W,'    ',',()#');
           CALL DWRIT(W,CW);
```

```
            END;
            END;
            END;
          ELSE IF SUBSTR(KAL_EL,3,2)='AN' THEN DO;
            IF SUBSTR(KAL_EL,5,1)¬='(' THEN DO;
            IF SUBSTR(KAL_EL,5,1)=' ' THEN DO;
            W=SUBSTR(KAL_EL,6,K1-6);
        W=TRANSLATE(W,'  ','.')#';
             CALL DWRIT(W,CW);
          END;
          END;
          END;
        ELSE DO;
          W=SUBSTR(KAL_EL,3,K1-3);
        W=TRANSLATE(W,'  ','.')#';
             CALL DWRIT(W,CW);
          END;
        END;
          IF SUBSTR(TI,LENGTH(TI),1)=';' THEN TI=SUBSTR(TI,1,LENGTH(TI)-1);
          L=1;
          DO WHILE(L¬=0);
        PARSE: L=INDEX(TI,';');
            IF L=0 THEN W=SUBSTR(TI,1,LENGTH(TI));
        ELSE W=SUBSTR(TI,1,L-1);
          IF SUBSTR(W,1,1)=' ' THEN DO;
          IF L¬=0 THEN DO;
            PITH=CW;
      WRITE FILE(NEPTUN) FROM(CHANG);
            TI=SUBSTR(TI,L+1);
            END;
          END;
          ELSE DO;
        CHECK='0'B; NCE='0'B; NCE='0'B;
        IF SUBSTR(W,1,1)='<' THEN DO;
            L=INDEX(SUBSTR(TI,2),'>');
            IF L=0 THEN DO;
              PITH=JET;
          L=-1;
              WRITE FILE(NEPTUN) FROM(CHANG);
              END;
            ELSE DO;
              W3=SUBSTR(TI,2,L-1);
          N1=INDEX(W3,'$');
              IF N1=0 THEN DO;
          N1=INDEX(W3,'+');
        IF N1=0 THEN DO;
        N1=INDEX(W3,'-');
        NCE='1'B;
        END;
      ELSE NCE='1'B;
                END;
        ELSE CHECK='1'B;
              IF N1=0 THEN DO;
                PITH=JET;
                WRITE FILE(NEPTUN) FROM(CHANG);
                END;
              ELSE DO;
                STR1=SUBSTR(W3,1,N1-1);
                STR2=SUBSTR(W3,N1+1,L-N1-1);
          K1=1;
```

```
        DO WHILE(K1¬=0);
          K1=INDEX(STR1,':');
        IF K1=0 THEN STORE1=SUBSTR(STR1,1);
        ELSE STORE1=SUBSTR(STR1,1,K1-1);
          K2=1;
        STRS2=STR2;
        DO WHILE (K2¬=0);
          K2=INDEX(STRS2,':');
          IF K2=0 THEN STORE2=SUBSTR(STRS2,1);
        ELSE STORE2=SUBSTR(STRS2,1,K2-1);
        IF NCE='1'B THEN W=STORE1||'-'||STORE2;
        ELSE IF CHECK='1'B THEN W=STORE1||'#'||STORE2;
          ELSE W=STORE1||' '||STORE2;
        CALL DWRIT(W,CW);
          W=TRANSLATE(W,' ','#');
        W=STORE2||'-'||STORE1;
        IF NCE¬='1'B THEN DO;
        IF CHECK='0'B THEN DO;
        CALL DWRIT(W,CW);
          END;
        END;
        IF K2¬=0 THEN STRS2=SUBSTR(STRS2,K2+1);
        END;
          IF K1¬=0 THEN STR1=SUBSTR(STR1,K1+1);
        END;
        END;
        END;
          IF L=LENGTH(TI)-1 THEN L=0;
          ELSE L=L+1;
        TI=SUBSTR(TI,L+1);
        END;
          ELSE DO;
              CALL DWRIT(W,CW);
            DWRIT: PROCEDURE(W,CW);
      DCL (W,W2,W3) CHAR(95) VAR,BIN CHAR(95) VAR,
          (LO,L2) FIXED BINARY, CW CHAR(132) VAR;
       L2=INDEX(W,'#');
          DO WHILE(L2¬=0);
          W2=W;
          BIN=SUBSTR(W2,L2+1);
          DO I=LENGTH(BIN) TO 1 BY -1;
           IF SUBSTR(BIN,I,1)='#' THEN DO;
           W3=SUBSTR(W2,1,L2-1)||' '||SUBSTR(BIN,I+1,LENGTH(BIN)-I)||
             '-'||SUBSTR(BIN,1,I-1);
           W3=TRANSLATE(W3,' ','#');
           CALL PWRIT(W3,CW);
           W3=SUBSTR(BIN,I+1,LENGTH(BIN)-I)||'-'||SUBSTR(BIN,1,I-1)||
             '-'||SUBSTR(W2,1,L2-1);
           W3=TRANSLATE(W3,' ','#');
           CALL PWRIT(W3,CW);
            I=1;
           END;
          END;
            BIN=W;
          W=TRANSLATE(W,' ','#');
          DO WHILE(L2¬=0);
            W2=SUBSTR(W,L2+1)||'-'||SUBSTR(W,1,L2-1);
        CALL PWRIT(W2,CW);
          LO=INDEX(SUBSTR(BIN,L2+1),'#');
          IF LO=0 THEN L2=0;
```

```
            ELSE L2=L2+LO;
        END;
          END;
         CALL PWRIT(W,CW);
            RETURN;
            END OWRIT;
        IF L¬=0 THEN TI=SUBSTR(TI,L+1);
          END;
          END;
        END;
          END FETCH;
      PWRIT: PROCEDURE(W,CW);
        DCL W CHAR(80) VAR,CW CHAR(132) VAR,VARW CHAR(38),
          WORD CHAR(132);
          VARW=W;
       IF SUBSTR(VARW,1,6)¬='INDEX-' & SUBSTR(VARW,1,9)¬='HANDBOOK-' THEN DO;
          WORD=(3)' '||VARW||(2)' '||CW;
          WRITE FILE(SORTIN) FROM(WORD);
         END;
          RETURN;
         END PWRIT;
      NEILS: XKA=N;
          PITH='  NUMBER OF RECORDS= '||XKA;
          WRITE FILE(NEPTON) FROM(CHANG);
       END KWOC;
```